数理解析研究所講究録
第1829巻 2013年 19-22

F-Divergenceに関連する問題について

神奈川大学・工学部　進藤　晋
Susumu Shindoh
Faculty of Engineering, Kanagawa University

1　はじめに

確率・統計や情報理論で現れるdivergence (relative entropy, cross entropy) は，2つの確率分布 p と q の間の違いの度合いを表す量として定義される [2]．一方，Csiszár[3] が定義したf-divergenceは，さまざまなdivergenceを特別な場合として含んでいる [12]．最適化の分野でも，f-divergenceを目的関数とする制約付き最適化（最大化）問題が考察されている [4, 5].

本論文の目的は，f-divergenceを定義する際に現れるadjointに関する性質をとりまとめ，さらにHiriart-Urrutyらが考察した関数方程式 [9] について解説することである．

2　Adjoint

$(0,\infty)$ で定義された凸関数のクラスを $\mathcal{F}$ とする．$f\in\mathcal{F}$ に対して，

$$f(0)=\lim_{t\downarrow 0}f(t)$$

が存在することに注意する．

$f\in\mathcal{F}$ に対して，f のadjointを以下のように定義する [1] ([11] では，*-adjointとよばれている).

定義 1 $f\in\mathcal{F}$ とする．このとき，f の **adjoint** f^{*} を

$$f^{*}(t)=tf(\frac{1}{t})$$

で定義する．

f のadjoint f^{*} は次の性質をもつ [1].

定理 2 $f\in\mathcal{F}$ に対して，

(1) $(f^{*})^{*}=f$

(2) f^{*} は凸関数

(3) $f(1)=0$ ならば，$f^{*}(1)=0$

(4) $f^*(0) = \lim_{t\to\infty} \frac{f(t)}{t}$

特に，2番目の性質が重要である．凸性の証明は，[8] にあるが，直接的な証明は見当たらないので述べておく．
(証明) $t_1, t_2 > 0, 0 < \alpha, \beta < 1, \alpha + \beta = 1$ に対して，

$$
\begin{aligned}
f^*(\alpha t_1 + \beta t_2) &= (\alpha t_1 + \beta t_2) f(\frac{1}{\alpha t_1 + \beta t_2}) \\
&= (\alpha t_1 + \beta t_2) f(\frac{\alpha + \beta}{\alpha t_1 + \beta t_2}) \\
&= (\alpha t_1 + \beta t_2) f(\frac{\alpha t_1}{(\alpha t_1 + \beta t_2) t_1} + \frac{\beta t_2}{(\alpha t_1 + \beta t_2) t_2}) \\
&\leq (\alpha t_1 + \beta t_2) \{ \frac{\alpha t_1}{\alpha t_1 + \beta t_2} f(\frac{1}{t_1}) + \frac{\beta t_2}{\alpha t_1 + \beta t_2} f(\frac{1}{t_2}) \} \\
&= \alpha t_1 f(\frac{1}{t_1}) + \beta t_2 f(\frac{1}{t_2}) \\
&= \alpha f^*(t_1) + \beta f^*(t_2)
\end{aligned}
$$

上の不等式の部分は，f の凸性を用いている．□

例 3

1. $f(t) = t \log t$ のとき，$f^*(t) = -\log t$

2. $f(t) = (1 - \sqrt{t})^2$ のとき，$f^*(t) = f(t)$

3 F-divergence

f-divergence の定義からはじめる．

定義 4 $f(1) = 0$ を満たす凸関数 $f \in \mathcal{F}$ に対して，$I_f : R_+^n \times R_+^n \to R$ を

$$I_f(p, q) = \sum_{i=1}^{n} q_i f(\frac{p_i}{q_i})$$

で定義する．I_f を **f-divergence** という．ここで，$R_+^n = \{p = (p_1, \ldots, p_n) : p_i \geq 0 \ for \ all \ i\}$.

上の定義で，$0f(\frac{0}{0}) = 0$，$0f(\frac{a}{0}) = \lim_{\epsilon\to 0} \epsilon f(\frac{a}{\epsilon})$ $(a > 0)$ と考えることにする．また，変数が確率密度関数のときは，I_f は積分で定義される．

$f(t) = t \log t$ とすると，I_f は Kullback-Leibler divergence，$f(t) = |t - 1|$ とすると，I_f は変動距離 (variational distance)，$f(t) = (1 - \sqrt{t})^2$ とすると，I_f は Hellinger 距離（の2乗）となる [1, 11, 12].

f-divergence I_f の性質をまとめておく（証明は [1, 11] 等参照）．

命題 5

(1) I_f は変数 p, q の凸関数である．

(2) $I_f(p,q) = I_{f^*}(q,p)$

(3) $f(1)=0$ を満たす $f \in \mathcal{F}$ が $t=1$ で狭義凸，すなわち，$t=1$ の近傍で f が線形とならないとき，

$$I_f(p,q) = 0 \Longleftrightarrow p = q$$

4 関数方程式

Hiriart-Urruty と Martinez-Legaz[9] は，情報理論にしばしば現れる adjoint と f-divergence の対称性に着目し，以下の問題を提起・考察した．

問題 6 関数方程式

$$f^* = f \tag{1}$$

を満たす凸関数を明らかにせよ．

関数方程式 (1) を満たす凸関数としては，上記の例 3 の $f(t) = (1-\sqrt{t})^2$ のほかに，$f(t) = |t-1|$, $f(t) = \sqrt{t^2+1}$ 等がある．

凸関数の性質から，以下の命題が成り立つ．

命題 7 *f を関数方程式 (1) および $f(1)=0$ を満たす凸関数とするとき，$f \geq 0$．したがって，すべての $(p,q) \in R^n_+ \times R^n_+$ に対して，$I_f(p,q) \geq 0$．*

Hiriart-Urruty らは，凸解析の手法を用いて，凸関数 f に付随する閉凸集合 C_f がある性質を満たすときに限り，凸関数 f が上の関数方程式 (1) を満たすことを示した [9]．ここで，C_f は以下のような集合である．

$$C_f = \{(x,y) \in R^2 : \text{すべての } s>0, t>0 \text{ に対して}, xs + yt \leq tf(\frac{s}{t})\}$$

5 おわりに

最後に，希望も含めて何点か補足しておく．

- Hiriart-Urruty らは，関数方程式 (1) を満たす凸関数を具体的に求めているわけではない．したがって，(1) を満たす凸関数あるいは凸関数のクラスを知りたい．
- 関数方程式 (1) を満たす凸関数の性質を明らかにしたい．
- 凸関数の adjoint の性質をさらに理解したい．
- 関数方程式 (1) は、$f(1)=1$ の条件のもとで，positive linear operators の分野で頻繁に出現している [10, 7, 6]．この分野でも，関数方程式 (1) を満たす解を知ることは有用であると考えられる．

[謝辞] 本研究は，科研費補助金（基盤C，No.22510161）の支援を受けている．

参考文献

[1] Ben-tal, A. et al.: Certainty equivalents and information measures : duality and extremal principles, J. Math. Anal. Appl., 157, 211-236 (1991)

[2] Cover,T.M and J.A.Thomas,: Elements of Information Theory, Wiley (2006).

[3] Csiszár, I.: Information-type measures of difference of probability functions and indirect observations, Studia Sci. Math. Hungar, 2, 299-318 (1977)

[4] Csiszár, I. and F. Matus : On minimization of multivariate entropy functionals, Proceedings ITW 2009 IEEE, 96-100 (2009)

[5] Gilardoni, G.L. : On the mininmum f-divergence for given total variation, C.R.Acad. Sci. Paris, Ser. I 343, 763-766 (2006)

[6] Hansen, F. : Characterizations of symmetric monotone metrics on the state space of quantum systems, Quantum Information and Computation, Vol.6, No. 7, 597-605 (2006)

[7] 日高文雄, 柳研二郎 : ヒルベルト空間と線型作用素, 牧野書店 (1995)

[8] Hiriat-Urruty, J.-B. and C.Lemarechalm : Convex Analysis and Minimization Algorithms I, Springer (1996).

[9] Hiriart-Urruty, J.-B. and J.-E. Martinez-Legaz: Convex solutions of functional equation arising in information theory, J. Math. Anal. Appl., 328, 1309-1320 (2007).

[10] Kubo, F. and T. Ando : Means of positive linear operators, Math. Ann. 246, 205-224 (1980)

[11] Liese F. and I.Vajda : On divergences and informations in statistics and information theory, IEEE Transactions on Information theory, Vol.52, No.10, 4394-4412 (2006)

[12] 須鎗弘樹 : 複雑系のための基礎数理, 牧野書店 (2010)